DOCUMENT RESUME

ED 354 003 IR 054 379

AUTHOR Engle, Mary; Lynch, Clifford
TITLE Directory of Electronic Information Resources: A Feasibility Study. Technical Report No. 4.
INSTITUTION California Univ., Oakland. Div. of Library Automation.
REPORT NO ISBN-0-913248-08-8
PUB DATE 20 Jun 90
NOTE 27p.; For related reports, see IR 054 378-381.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Libraries; *Access to Information; Bibliographic Utilities; Change Strategies; Computer Software; *Databases; Feasibility Studies; Higher Education; Library Automation; *Library Networks; *Machine Readable Cataloging; *Online Catalogs
IDENTIFIERS Electronic Text; MARC; MELVYL; *University of California

ABSTRACT

This paper provides an overview of a University of California (UC) study of the issues involved in creating an online directory of electronic information resources available to the university community. For the purposes of the study, electronic information resources include bibliographic and other databases available at or through UC libraries and campus computer centers, and databases and data files maintained in departments. Such a directory would bring together information about computer files that has not been available from one source. It was found that the project would be feasible if the directory could be implemented in three phases: (1) design of the record structure, user interface, and means of access; (2) implementation and evaluation of a prototype; and (3) refining the user interface, indexing additional fields, and adding newly created records. Six development issues (some of them complex) that would need to be resolved before implementation were identified: the forms of electronic information to be included; the scope of the database, including external databases; availability of satisfactory descriptive information about items to be included; and a machine readable catalog record for each item. It is noted that, in addition to existing UC library records and other on-campus sources of electronic resources, there are nine bibliographic utilities and commercial vendors which produce "databases of databases" that could be included in the UC directory; other resources are available to provide descriptive detail for items in the directory. It was concluded that the directory should exist as a separate database searchable by the MELVYL user interface. Further discussion about Machine Readable Computer Files (MRCF) and a list of the MARC Bibliographic Format for MRCF are appended. (KRN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED354003

Technical Report No. 4

DIRECTORY OF ELECTRONIC INFORMATION RESOURCES
A FEASIBILITY STUDY

July 20, 1990

Prepared by

Mary Engle
Clifford Lynch

Division of Library Automation
University of California
Office of the President
300 Lakeside Drive, Floor 8
Oakland, CA 94612-3550

IR054379

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard West

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

©1990 The Regents of the University of California

MELVYL is a registered trademark of The Regents of the University of California.

Portions of this document may be reprinted or adapted without permission for academic nonprofit purposes, providing the material is accurately quoted and the source duly credited.

ISBN: 0-913248-08-8

CONTENTS

1. Background 1

2. Project Overview 2

3. Summary of Findings 2

4.0 Feasibility of the Project 3

5.0 Discussion of the Development Issues 5

- 5.1 Types of Electronic Information 5
- 5.2 The Scope of the Database 6
- 5.3 Data Sources 8
  - 5.3.1 Cataloging of University of California Resources 8
  - 5.3.2 Non-UC Sources 9
  - 5.3.3 Linking Additional Information 13
- 5.4 Content of the Directory Entries 15
- 5.5 Means of Accessing and Searching the Directory 15
- 5.6 Update Mechanism—Maintenance of Records over Time 16

6. Proposed Action 16

Appendix A Machine-Readable Computer Files (MRCFs): Further Discussion

Appendix B MARC Bibliographic Format for Machine-Readable Computer Files (MCRFs)

Technical Report No. 4

DIRECTORY OF ELECTRONIC INFORMATION RESOURCES A FEASIBILITY STUDY

1. BACKGROUND

Information in electronic form represents a growing resource of significant importance to the University of California community, and one that is increasingly vital to research and instruction. A great deal of money and effort is being expended in creating, acquiring, mounting, and maintaining information resources throughout the University. However, locating and identifying information available in electronic form is a major difficulty within the UC community: There is no single source for information on electronic data files, nor is there currently a single source for this type of information on any individual UC campus.

In its final report to Library Council (March 8, 1989), the UC Electronic Information Review Committee (EIRC), chaired by Professor David Phillips of UC San Diego, recommended that the University of California develop and mount a database describing available electronic information resources for access by the UC community. The Committee recommended that the Office of the President's Division of Library Automation (DLA) coordinate the development of this database and mount it centrally, to be accessible throughout the UC system via the MELVYL® system. The proposed online directory would be the primary source of information for the UC community on the availability and accessibility of electronic data resources for the University of California community.

This paper provides a project overview, examines the issues involved in creating an online directory of electronic information resources, and proposes a multi-phased approach to the creation of the directory.

Clarifying the Terminology

The Committee defined the term "electronic information resources" as liberally interpreted to include bibliographic or other databases or electronic resources available at or through UC libraries or campus computer centers, and databases or data files maintained in departments. Thus, items of interest represent a broad range of machine-readable materials, including small or large UC-owned or public-domain databases, databases accessible by telephone dial-up or through the national network, software programs, numeric and statistical files, raw data files, machine-readable lists, and textual information. The terms "database" and "directory" are used interchangeably where they describe the product of this project.

2. PROJECT OVERVIEW

The goal of this project is to assist members of the UC community in identifying, locating, and exploiting the broad range of electronic information resources available to them, including those within the UC system, nationally and perhaps internationally. DLA's objective is to deliver an online database of citations to existing electronic resources of interest to the UC community.

The directory would bring together disparate information sources to which there has traditionally been inconsistent or, in many cases, no previous bibliographic access. It would tell the user that certain electronic information sources exist, provide information *about the information sources*, and help the user determine how to access them.

The directory would be mounted as a database accessible via the MELVYL system, and would contain entries which include such information as the type of resource, the producer or source of the information, where to obtain access to it, a physical description of the medium, and as much detail as possible about the content. In this sense, it goes beyond the typical cataloging citation and also describes materials not covered by normal cataloging.

The directory would be most useful to UC if it contained citations to individual materials as well as collection-level references, such as the Census collection, the U.S. Naval Observatory Electronic Data Tapes, or the U.S. Bureau of Labor Statistics electronic data. It would include citations to individual data files or pieces of software created at UC, to commercial databases, and to the collections of statistical and scientific information available from private, government, and other university sources. Electronic resources available on local, regional, and national computer networks are becoming an increasingly vital part of the university information sources for research and instruction; inclusion of these types of materials would also enhance the value of the directory as a resource.

3. SUMMARY OF FINDINGS

The directory is unique because it would bring together descriptive information about computer files that has not previously been made available from one source, and because it would go beyond the traditional concept of a library cataloging only its own holdings. Data sources are discussed in Section 5.3 and supplemental information in Section 5.3.3. The multiplicity of dissimilar data sources will require a considerable programming effort by DLA staff to convert the data to a common format, consolidate duplicates, and load the data into the directory's database.

This study finds that the project is feasible if the directory can be implemented in phases. The first phase would consist of the design of the record structure, user interface, and means of access.

The second phase would involve the implementation of a prototype with a test period to gain access to the database and experience with it. The directory would be bulk loaded with existing machine-readable records describing electronic resources, both from UC sources (existing MELVYL catalog records for computer files) and non-UC sources such as a commercial directory of databases (Section 5.3.2).

The third phase would involve refining the user interface, implementing changes such as indexing additional fields, and adding newly created cataloging records for UC holdings. Future records for UC holdings would be added through the normal MELVYL catalog input streams (Sections 5.3.1 and 5.6).

Development issues include:

- Identifying the types of electronic resources to be included in the directory;
- Defining the scope of the database and the data sources;
- Defining the data elements that constitute a record;
- Establishing the means of accessing the directory;
- Defining its user interface; and
- Determining how the directory will be maintained over time.

This study proposes that an advisory task force be created to work with DLA to determine the nature and scope of the directory and mechanisms for data collection, record creation, and update (Section 6).

4. FEASIBILITY OF THE PROJECT

The project to mount a Directory of Electronic Resources on the MELVYL system is feasible given the breadth and depth of existing resources:

- The proven expertise of the staff of the Division of Library Automation in mounting databases, both from internal UC and external input sources,
- The ability to integrate the directory into and manipulate it within the MELVYL system,
- The multiplicity of possible input sources described below, including summary information on electronic resources in machine-readable form available both commercially and in the public domain, and

- Campus efforts, both planned and already initiated, to identify and create access to electronic information.

Since a number of the issues involve acquiring data from multiple sources and developing mechanisms for record creation and maintenance over time, it would be necessary to implement in stages the elements that would eventually comprise the full directory. It would be feasible in incremental phases, beginning with the design phase in the next fiscal year, if we were able to exploit existing summary information on resources used by or of interest to the UC community, while UC campuses develop strategies to collect and create catalog records for UC holdings.

Prototyping will be a necessary second phase to allow us to examine the complexity of mapping records from multiple external sources to the MARC format, and combining different types of data (some of which are extensions to records, rather than individual records themselves). In prototype, we will have some ability to do manual manipulation of records in the process of converting and loading records into the database. This experience will help determine what should be done manually and what automated tools we may wish to add later.

The bulk loading of existing UC MELVYL catalog records and one or more commercial or public-domain compilations of such information, mapped to the MARC format with appropriate extensions, would be a strong beginning. Other UC records would be added as they are identified and cataloged by UC libraries, marked for inclusion in the directory, and input through the normal MELVYL catalog input streams.

The third phase would involve refining the user interface based on experience with the database and feedback from the MELVYL System User Services Group and other UC librarians. Such experience may suggest additional fields to index or other changes to the final production version. The directory would then be mounted as a database accessible within the MELVYL system.

This paper examines the following development issues in more detail to provide a foundation upon which to build such a phased approach:

- Types of electronic information
- Scope of the proposed directory of information resources
- Data sources for the directory
- Content of directory entries (ie, data elements)
- Means of accessing and searching the directory
- Maintenance of the directory over time

This list indicates that there are a number of complex issues to be resolved, particularly in regard to the creation and maintenance of records for UC-owned resources.

We can begin immediately, however, to pull together information on databases and data files of external origin that are accessible by the UC community, and are thus within the Committee's recommended scope, as well as the information we now have on locally held computer files and databases. Later, as records for UC holdings are created incrementally, we can add them to the existing database.

5. DISCUSSION OF THE DEVELOPMENT ISSUES

5.1 Types of Electronic Information

Machine-readable resources take many forms. The following types of materials are examples of what is currently available:

- Textual Resources — Eye-readable textual information such as bibliographic citations, full text of articles, facts from fact databases, or simple lists of information.

- Numeric or Quantitative Resources — Tables or complex arrays of numeric information, such as census data or statistical information systems.

- Image or Graphic Resources — Databases of nonbibliographic materials such as digitized slides, LANDSAT images, maps, and artwork. (The directory could be an extremely valuable information source for just such new and little-known materials.)

- Software Resources — Computer programs that organize or manipulate data, perform useful tasks, function as operating systems or utilities, teach (such as CAI), or inform (such as expert systems).

Individual members of each of these categories may differ widely in organization and sophistication, from those having a highly structured organization with descriptive documentation (such as DIALOG databases) to "raw data" files that are simple, undocumented lists of text or numbers. They may also differ in a number of other ways—for example, resources from which the user can extract and utilize data directly vs. those requiring sophisticated programs to create, manipulate, or format the result desired by the database user. We should assist the user by indicating in the record the degree of organization and the structure of the item described.

Other forms of machine-readable resources may form boundary cases for consideration. Are the less formal sources of online technical information, such as bulletin boards, appropriate resources for the directory? For example, there are over 20 Bulletin Board

Services (BBSs) run by government agencies. There may be other, less well defined areas that come to light when the survey of current resources is complete. As we gain some experience with the various types of materials, we may need to focus or expand our efforts.

5.2 The Scope of the Database

The Committee's general recommendation is to create a database of broad scope that would satisfy a wide variety of requests. The scope of any directory is limited, however, by certain objective criteria. The Committee suggested, as initial criteria for inclusion in the directory, that files either be available through UC libraries, computer centers, or departments, to at least some members of the UC community, and be reasonably maintained.

This is a broad recommendation since both UC libraries and computer centers access electronic data resources nationwide from both public and private sources, and, in fact, end users can access these resources over the network. Each year the number of new databases dramatically increases. The directory would best serve the UC community by providing access to the broadest array of information on external databases.

The following discussion elaborates on these general guidelines and others.

- Accessibility — In general, the resource must be available to some reasonably broad community. Access need not be free for a resource to be included, nor must the resource itself offer any means to manipulate the data. For example, an astronomer who created a data file detailing radio sources within a radius of *n* degrees of the sky may not offer software to manipulate the data, but the file may still be of interest to other UC researchers prepared to obtain their own tools to manipulate the file. One implication of this is that a resource provider for census data, for example, need not make available computing resources to analyze the data. A user of the file would transfer data to some other machine on which he or she had obtained software and cycles to manipulate the data.

- File Maintenance — The Committee stated that files in the database must be reasonably maintained. We will probably need to further define reasonable maintenance, but this would probably include some commitment to update the file (where appropriate), to provide computer cycles to make it available via the network, to keep the machine housing the data available (where relevant), to share the data, to document the data, and to answer questions.

  It is important to distinguish between static and dynamic files, perhaps as a data element in the file's entry in the directory. Whereas bibliographic databases are normally dynamically maintained by additions, revisions, and deletions, image

archives or scientific data may remain static. In such cases, a commitment to ensure that the computer supporting the image database remains available to the network might constitute reasonable maintenance.

- Minimum Descriptive Information — We need to have enough information to describe an item at least minimally and to determine its location and accessibility. While descriptions for directory entries may vary dramatically depending on the source or type of item, the descriptive record should contain a requisite set of data elements. The following data elements are a preliminary guide for providing the minimum amount of information necessary.

Directory Field	UC Brief/MARC Standard Field
Title or name of the resource	Title statement (245 $a)
Source – Responsible party (where to obtain access)	Name of publisher, distributor, etc. (260 $b)
Date (of last modification?)	Date of publication, distribution (260 $c)
Source – Place	Place of publication, distribution (260 $a)
Type of computer file (e.g., database, computer program, text file)	Physical Description (300 $a – $e) and 008 field, position 26
Medium	Title statement—medium (245 $h)
Version	Volume Designation (362)
Packaging Method	
Content	Contents (505 or 520)
Restrictions on access	Restrictions on Access Note (506 $a)
System requirements	Technical Details Note (system requirement – 538 $a)
Notes*	General Notes (500) or Summary or Abstract (520)

* For computer files, the Notes field often carries information important to accessing many esoteric records.

- Machine-Readable Record — We must have a machine-readable catalog record for each item. Creation of records describing campus resources is outside the scope of DLA's immediate responsibility, though campuses may create them in support of this project. DLA could cooperate with campus groups in seeking grant funding that covers both UC campus resources and special campus projects for descriptive cataloging of external resources such as network-available resources, government collections of electronic data, and dial-in access services. For these types of resources that are available to the University as a whole, DLA may be an appropriate lead organization, although, for the actual cataloging, DLA would either partner in grants or fund activities at a UC campus or other organization.

For certain types of external resources, collaborating with interinstitutional consortia may be an appropriate way of obtaining catalog records, especially if different institutions have special subject expertise.

There are at least three ways in which the record for an item can be included in the database:

1. It is created by a UC library or department. This can be accomplished through normal cataloging channels for UC holdings, or through special subcontracted, DLA- or externally funded projects where actual cataloging would take place at a UC campus or other organization.

2. It is transferred from a bibliographic utility (e.g., RLIN, OCLC), and perhaps enhanced with additional data.

3. It is transferred from another system or a file that has been obtained by UC (for example, the Cuadra Directory of Online Databases). We can reasonably assume that it will be necessary to reformat or enhance data incoming from other bulk files, such as a catalog of databases. DLA can accomplish this working directly with the vendor.

5.3 Data Sources

5.3.1 Cataloging of University of California Resources

Existing Library Records

UC's references to its own holdings of machine-readable files are a small but growing body of bibliographic records. Catalog records for electronic data resources are often referred to as machine-readable computer files, or MRCFs (formerly known as machine-readable data files, or MRDFs). The existing bibliographic records for MRCFs in the MELVYL catalog can be duplicated in the directory as one of its initial bulk-loaded files.

The MELVYL catalog currently holds approximately 667 records for computer files, a number that continues to grow at a steady pace. MRCFs in the catalog have increased by more than 25% over the past four months alone. Except in a few cases, UC libraries only catalog what they have, not what they have access to. Thus, UC catalog records for machine-readable computer files will never be a complete data source for the directory of databases. Software programs constitute a major portion of the existing MELVYL catalog records for machine-readable resources already cataloged by UC libraries.

At least initially, the focus should be on databases, including software only where relevant to access or use of an electronic resource, or where it stands alone as a network service. Two exceptions are major software servers that are, effectively, databases of software (e.g., at the University of Michigan), or possibly, major program libraries. The objective is to avoid including 50 records for library holdings of DOS 3.1.

If we really want to cover software, there are a number of software directories that we could license. If we include software, we could also include reviews of software available in full text form, licensed from periodicals such as *Byte*.

Some of the issues surrounding creation and management of MRCFs are discussed in Appendix A. The set of fields comprising the MARC record for MRCFs is included as Appendix B.

Campus Efforts

UC campuses have begun efforts to bring machine-readable data under bibliographic control, though these efforts are relatively new; UCLA is the first campus to undertake a survey of locally held electronic materials. Two campuses have full-time database librarians. Libraries at Riverside and Berkeley have media centers that create machine-readable records for their holdings. Several campuses have librarians whose focus is bibliographic control of machine-readable items. The meeting of Data Archivists in Berkeley in September 1989 demonstrated that virtually all campuses have active people and machine resources devoted to providing both bibliographic and direct user access to machine-readable computer files.

Since we lack information on the majority of machine-readable resources held by UC, the Electronic Information Review Committee recommended that the Office of the President, through the Office of Library Affairs, undertake a University-wide survey of electronic information resources available to the UC community. Staff at the Office of Library Affairs are studying the feasibility of such a survey. This survey would identify resources to be included in the directory of databases. There is much work to be done before a true representation of UC holdings will be available.

5.3.2 Non-UC Sources

There are a number of sources of information on electronic resources. For example, several commercial vendors produce "databases of databases," and national bibliographic utilities, such as OCLC, RLIN, and WLN, hold MARC records for MRCFs created by libraries other than those in the UC system. The following discussion of some of the major sources assumes appropriate license agreements could be negotiated for use of these sources.

ICPSR Files

The Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan provides extensive data files on poll and census information, currently collected and managed on both the Berkeley and UCLA campuses. UC campuses are members of the ICPSR consortium, thus having access to this information.

Records describing the ICPSR electronic data files will be an important addition to the directory at the collection level. Further, the University of Michigan creates full MARC catalog records in the RLIN database for each data file, with extensive summaries of contents in the 520 field. Users of these data abstracts have long needed keyword access to information in the Notes fields, which we could provide via the directory.

Commercial Directories of Databases

Online databases are appearing at a dramatic rate—approximately two per day. CD-ROM databases will soon match this rate and surpass it. There are currently three commercial databases that describe databases in machine-readable form: the Cuadra Associate's *Directory of Online Databases*, the *Gale Research Online Database of Databases* (newly available as the DIALOG Database of Databases, File 230), and Knowledge Industry Publications' *Directory of Databases*. The databases listed in these directories are available to the UC community.

The Cuadra Directory of Online Databases

Cuadra Associates produces an extensive summary of databases available in machine-readable form for $4,000 annually, including quarterly update tapes. The directory currently contains over 4,000 listings of databases generally available to a broad audience through a variety of large and small commercial vendors and government agencies.

The directory includes highly specialized and lesser known databases, available both nationally and internationally. It provides information on the type of database, subject area, producer, online access, content, coverage, time span, and update cycle. It currently does not contain information on CD-ROM databases, but the publisher plans to add this in the near future. We already have specification sheets on record layout for this database.

Cuadra has announced the CD-ROM information as a second publication—*The Directory of Portable Databases*, including databases on CD-ROM, Bernoulli cartridges, floppy disks, or magnetic tape. The *Directory* exists in both print and online versions, describing over 600 databases.

The record structure for the *Directory of Portable Databases* is slightly different from the original online database, due in large part to the difference in medium. There are

important elements of both search software and hardware that must be noted for CD-ROM databases that are not necessary parts of the online database records. A special field in each database indicates whether the database is online or portable.

The Cuadra databases have the following file sizes:

Directory of Online Databases — approximately 15 MB (with Cuadra's indexing).

Directory of Portable Databases — 5 MB and growing rapidly.

Cuadra's production schedule suggests that they will either have the databases merged by the time we would be ready to load, or that an annual subscription for each will be available (if the databases do not merge well). Cuadra is attempting to keep the price for the merged databases at or close to the $4,000 quoted price.

The Gale Online Database of Databases

DIALOG uses the Online Database of Databases published by Gale Research. Gale's product includes information on CD-ROM databases but is not available directly from Gale Research. Gale does not foresee its availability to end licensees like UC in the next year.

Knowledge Industry Publications Database Directory

BRS uses the Knowledge Industry Publications (KIP) Directory as its database of databases, covering all types of databases, with a focus on those produced in the U.S. and Canada. The five sections of the print version of the directory derive from two separate databases—the list of approximately 2,500 databases (arranged numerically by record number rather than alphabetically) and the vendor/producer index with 1,200–1,500 entries, including names, addresses, and pricing information. There is also a subject index, although it is limited to 60 headings. It does not contain information on CD-ROM databases.

KIP will make the file available to us for a fee of $2,500. Though the company produces the print version once a year plus a semiannual update, this program of distribution provides for only one tape per year with no updates. The $2,500 will be an annual fee if we wish to purchase a new copy of the database for updates.

Each new printing cycle adds about 200 new databases. It was last updated in the end of July 1989. The file size is approximately 47 MB. KIP can provide us the edited version produced as a flat ASCII file before the typesetting codes are entered. We have specification sheets on record layout for this database.

Resources from Commercial Bibliographic Utilities

Records representing electronic data files total over 21,000 in the OCLC database. The RLIN database holds 5,583 MRCF records. Search capabilities in each of these systems limit the ability to determine how many of these records the University of California holds. It may be possible to acquire the entire MRCF collection from one or both utilities, but would probably not be useful in its entirety since many of the records represent software holdings of other libraries. Alternatively, we may wish to acquire special collections like the ICPSR files from these sources.

For UC data, the bibliographic utilities can serve as they do for retrospective conversion projects to identify UC-held records with existing cataloging.

Internet Resource Directories

Summary guides to resources available on national computer networks are growing both in number and quality. Several guides currently exist; examples include the NSFNET *Internet Resource Guide* and the *Internet-Accessible Library and Databases Catalog*, available from the CERFnet Network Information Center.

The *Internet Resource Guide*, published electronically by the NSF Network Service Center, is itself a growing summary of electronic resources available on the Internet, including information on supercomputing facilities, library catalogs, other computer networks, etc. We should provide access to this information, perhaps both in the manner that we do for the *DLA Bulletin*—as a pageable document displayed from the catalog—and by including the *Guide*'s entries in the directory (with appropriate copyright authorization from the NNSC). The *Internet-Accessible Library and Database Catalog* focuses on library catalogs and information databases, overlapping only slightly with the NNSC's *Internet Resource Guide*.

Others directories and listings of resources have been proposed for EDUCOM and ALA's Library and Information Technology (LITA) Division. The existing guides are freely distributed and currently maintained, making them valuable additions to UC's directory.

Electronic Journals and Discussions

Both the electronic journals and Internet discussion mailing lists (moderated or unmoderated) are examples of electronic resources that would warrant entries in the directory.

Many of the electronic journals now beginning to be disseminated on national computer networks have editorial policies similar to those of their print counterparts in a variety of professional areas. Internet lists are electronic discussions of technical and nontechnical issues conducted by electronic mail over the Internet. Participants subscribe via a

central service, and lists often have a moderator who manages the information flow and content. These can be viewed as a sort of continuously published journal not covered by abstracting and indexing services.

The MELVYL catalog currently provides access to the text of locally produced electronic journals such as the *DLA Bulletin* and the *Mynd of the MELVYL® Catalog* (*MOM*), and in the future, is likely to be the primary means of locating other journal articles available in electronic form through abstracting and indexing databases such as MEDLINE®.

One approach for electronic journals and lists is to create directory entries as well as CALLS records for them, and provide public access to selected ones via the MELVYL catalog in a manner similar to access to the *DLA Bulletin* and *MOM*.

Directories of Federal Data Repositories

U.S. Government data in electronic form is abundant, though public awareness of it may not be. Examples include the U.S. Department of Labor Statistics Electronic Data Distribution program and over 20 bulletin board services, such as the Department of Commerce, Bureau of Economic Analysis, 24-hour data line offering the latest economic data from government agencies. For these types of government information resources available in electronic form, the Directory is an ideal current awareness vehicle.

5.3.3 Linking Additional Information

Other sources can provide supplemental information that add descriptive detail to citations in the directory and substance to cryptic records, thereby greatly increasing the value of records in the directory.

DIALOG Blue Sheets

The *Blue Sheets* are files of information describing the search and output capabilities for each database available through DIALOG. Available in machine-readable form and online as DIALOG File 415, the *Blue Sheets* contain information on the fields indexed, the syntax of search statements, and output formats. Linked to the bibliographic record describing that database, the *Blue Sheets* become a unique information source for the user to determine whether or not to access a particular database and how to enter a search. DIALOG considers this project a marketing tool, and is making the Blue Sheet data available to us at a nominal cost.

Other Types of Information

The following types of information similarly enhance the utility of records describing databases:

- The CONSER serials records in the OCLC database were supplemented several years ago with a field indicating where the serials were abstracted and indexed. From the information in this field, we could derive a list of the journals indexed in a given abstracting and indexing database to assist users in evaluating the utility of a database. The new CALLS database has already set the stage for this.

- Informational screens that provide search formulation examples for a database with a complicated structure.

- Document delivery facilities, where available, that are relevant to a database (for example, the ERIC Document Reproduction Service (EDRS), and document delivery services for Chemical Abstracts, Mathfile, ISI databases, and UMI Dissertation Abstracts).

- Statistical databases such as census databases require users first to refer to code books describing positional data elements. (Adding online access to the information in the reference code books obviates the need for the print version of the code books and opens up the database to users in remote locations.)

  This is an example of the more general case of acquiring documentation *about* a resource in electronic form and making it available online, either by adding it to the directory's records, or from a remote server linked via the MELVYL catalog.

Programming Requirements for Multiple Input Sources

Merging dissimilar electronic records requires a considerable programming effort to convert them to a common format (MARC), consolidate duplicates, and add extensions. The consolidation of duplicates is a design goal, but we recognize that this will be difficult. The multiplicity of input sources described above represents an equal number of programming tasks since only records created by UC campus libraries will have a uniform format when they arrive.

This programming effort will place serious demands on DLA's programming and production staff, as well as the documentation staff, in presenting the diverse collection of resources to users. Implementation in phases will spread this effort over time, but most will likely occur within a single calendar year in phase two. Future changes to the data or data structure implemented by external vendors will require similar changes to programs created to load data into the database.

5.4 Content of the Directory Entries

The directory will be created and updated from a variety of sources. OCLC, RLIN, and campus library MRCF records are available in MARC format, but most other records are not. The commercial directories of databases consist of extremely brief records that we could map to the MARC format. (*Standard for Brief Machine-Readable Bibliographic Records for University of California Libraries*, available from DLA, defines the minimum data elements required for inclusion of a cataloging record in the MELVYL catalog.)

The database may also contain many different types of information linked to base records, much of which bears little relation to the data elements in the MRCF record format. Some records may even be composites from multiple sources. We need to determine what minimum set of data elements provides the user with enough information to be of value. The dozen or so data elements listed in Section 5.2 on scope of the database may be used as the basis for future work in collaboration with campus representatives.

Since only some of the records will be actual UC holdings, it will be necessary to include in the record either a holdings statement or some indication of how a user can gain access to the electronic data file.

Problems of Subject Control

It is likely that we will have many different input sources for directory records. The difficulties of controlling subject vocabularies will be amplified by the number of data sources and the wide range of subjects represented. Keyterm indexing additional fields—for example, the MRCF Notes fields or the brief textual description of databases in Cuadra—increases the accessibility of a record.

Subject access and vocabulary control are areas in which we will need further study, and are likely to be an ongoing problem.

5.5 Means of Accessing and Searching the Directory

The directory will exist as a separate database, searchable by the MELVYL user interface. With the SET DB command, the user may select the directory from a welcome screen, or switch to it at any time during the session. This approach is described in Mike Berger's paper "Integration of Multiple Databases into the MELVYL Catalog."

The directory would be mounted centrally at the Office of the President, with user access by the same methods used to reach the MELVYL catalog (i.e., by hard-wired terminals, network or telephone dial-up access).

Within the directory, the nature of its records requires special access points beyond the usual author, title, and subject indexing. The following indexed additional access points are recommended for machine-readable files:

- Source machine
- Type of computer file (database, computer program, etc.)
- Notes field
- Means of access

Limits by date and medium are necessary points of fine tuning.

5.6 Update Mechanism—Maintenance of Records over Time

Some of the information contained in this directory will be volatile as either the file's content or its means of access changes.

There will be two general categories of records in the database: UC-generated records and records bulk-loaded from files of external origin. UC-generated records are those created by UC librarians, describing files owned by members of the University community. External files are those collected from sources outside the University community, such as commercial databases and government census files.

Non-UC records are reasonably easy to maintain since the updated collective files can be reloaded annually or on other cycles. We will need to develop a mechanism to interact with the commercial information vendors to supplement and update the database on a regular cycle, such as a periodic reload of any commercial files that we have incorporated into the directory.

UC records will have to be maintained by campus libraries. Assuming that the ultimate responsibility for creating and updating records lies with campus libraries, DLA should simply be able to accept in the normal input stream records that update existing records.

6. PROPOSED ACTION

It will be necessary to resolve the issues discussed above in order to define more precisely the product that the University wants to deliver and the mechanism for developing and maintaining it. For previous projects, advisory groups have played an important role in providing design direction and feedback on development of the user interface. We propose to establish a similar group, consisting of DLA and Office of Library Affairs staff and campus representatives to define the following major aspects of the project:

1. Nature and scope of the database
    a. The scope of the directory
    b. The nature and form of its descriptive records

2. Mechanisms for data collection, record creation, and update.
    a. What is the best way to gather the University's data?
    b. Who should create directory records?
    c. How will the records be maintained?
    d. What data should be indexed?

In addition to its design and implementation advisory role, the group can advise on database identification, and record production and maintenance.

Prototyping is essential, and there is no precedent for this type of directory. Since the *Directory of Electronic Resources* is relatively small (at least in prototype), it could be housed on a workstation, allowing fast development.

This undertaking is of great national importance. Grant funding should be readily available. The type of file linkage we are proposing (e.g., CONSER abstracting and indexing information and DIALOG Blue Sheets) has not yet been done, so we will need optimal data on the degree of difficulty of the task. In parallel, we should seek funding for a prototype, convene the intercampus committee to interact with the prototype development, and continue consideration of broader issues.

Appendix A

Machine-Readable Computer Files (MRCFs)

Further Discussion

A subset of the MARC record describes machine-readable data files. Recently, it was renamed to Machine-Readable Computer Files (MRCFs). The standard for MRCFs includes fields for both monographic and serially issued MRCFs; the implementation of format integration in the 1990s combines these. Although MRCF records carry much relevant information, for the broad purposes outlined here they are limited in scope.

Currently, librarians seem split over the utility of the MRCF serials format. Many are using the standard MARC serials format instead, since it provides more relevant fields. The monographic version of the MRCF format seems universally accepted. The standard MARC record for MRCFs is attached as **Appendix B**.

Standard for Brief Machine-Readable Bibliographic Records for University of California Libraries, available from DLA, also includes a subset of fields that describe MRCFs.

Extending the Description

Both the MARC record and the UC minimum standard record formats lack fields for some critically useful data elements. For example, information can be included in records to address the questions:

- Through what commercial services is this database available?
- Where on campus can a mediated search of a database be done?
- If this database is on a server, what is its name?
- What journals are indexed or abstracted in the database?
- What database fields are searchable?

We should make serious efforts to extend the description of entries to include this type of reference information, as well as local reference (e.g., location) and holdings information.

The following are suggestions on the types of additional information that would be extremely useful to the UC community. There are undoubtedly other ways that we could extend the bibliographic data to more completely describe electronic data resources.

1. For databases organized hierarchically in a tree structure, make a list of the classification codes available online.

2. Provide an online brief guide to searching the database.

3. Indicate in a Notes field where one can get a mediated search of a database.

4. Indicate UC ownership or holdings of a database or datafile.

Collection-Level Cataloging

For certain departments and institutions that hold or produce large amounts of machine-readable data files, we may wish to provide only collection-level cataloging. For example, a campus astronomy department may hold hundreds of data files from the U.S. Naval Observatory. A single collection-level entry may suffice to indicate the existence of such a set of materials.

Format Integration

To the extent that we use the MARC standard MRCF format, we will need to provide for the impending implementation of format integration. The Format Integration proposal has been accepted and will become a revision to the US MARC formats in the 1990s. Format integration proposes a single bibliographic format with all data elements valid for any kind of material. It also provides for the description of seriality in addition to the primary material description. The Library of Congress will implement format integration in 1993; the date is setting the pace for national implementation of these revisions to the standard.

It appears that the changes imposed by format integration will improve the situation of MRCF cataloging. In general, some of the changes that should alleviate historical problems with MRCFs are:

- Extended validity—all fields will be valid for all materials.

- Additions to fields—for consistency or in cases where two fields were merged into one.

- Changes to names of fields—for clarification when the field was taken out of the context of the particular MARC format. For example, "File Characteristics" (tag 256) becomes "Computer File Characteristics" in the integrated format.

- The 006 field is a new field that carries fixed field information for secondary material characteristics of the item being described. Under Format A, an 006 field can be used to express seriality, for example.

Considering the lack of agreement over the MRCF serials format, the introduction of format integration will probably make the cataloging of MRCFs easier. Some fields that have been saved from obsolescence are particularly useful for computer files:

516 Type of file or data
522 Geographic coverage
556 Documentation
567 Methodology
582 Related computer files

Indexing

To provide the necessary access points to MRCFs, the following variable-length fields have been recommended by MRCF catalogers as important to index in addition to the basic Title, Author, and Subject fields:

MARC 036 Original study number
037 Stock number
211 Acronym or shortened title
214 Augmented title
753 Technical details for access
(Machine type, operating system, program language)

In addition to the fields above, entries should include information on such matters as

- Information on restricted access
- Special software needs
- Charges associated with the database
- Contact person or department

Resolution of these issues would be the domain of the recommended task force.

Appendix B

MARC Bibliographic Format for Machine-Readable Computer Files (MRCF)

(M = Mandatory; A = Mandatory if applicable; O = Optional)

Field	Notes	
Leader	M	All character positions are defined
Control	**Fields**	
001	M	Control Number
005	M	Date and Time of Latest Transaction
006	M	Linking Field
008	M	Fixed-Length Data Elements
Numbers	**and Codes**	
010	A	Library of Congress Control Number
015	O	National Bibliography Number
017	O	Copyright Registration Number
020	A	International Standard Book Number
022	A	International Standard Serial Number
035	O	System Control Number
036	O	Original Study Number
037	O	Stock Number
040	M	Cataloging Source
041	A	Language of Text Files
042	A	Authentication Code
043	A	Geographic Area Code
045	O	Chronological Coverage (see 523)
050	O	LC Call Number
051	A	LC Copy, Issue, Offprint Statement
052	O	Geographical Class Codes
055	O	Call Numbers/Class Numbers Assigned in Canada
060	O	NLM Call Number
066	A	Character Sets Present
070	O	NAL Call Number
072	O	Subject Category Code
074	O	GPO Item Number
080	O	UDC Classification Number
082	O	Dewey Decimal Classification Number
086	A	Government Document Classification Number
09x	O	Local Call Number

100	A	Main Entry – Personal Name
110	A	Main Entry – Corporate Name
111	A	Main Entry – Meeting Name
130	A	Main Entry – Uniform Title
211	O	Acronym or Popular Title
214	O	Augmented Title
240	O	Uniform Title
242	O	Translation of Title
243	O	Collective Uniform Title
245	M	Main Entry – Title
250	A	Edition Statement
256	A	File Characteristics
260	A	Publication, Distributor (Imprint)
263	O	Projected Publication Date
265	O	Source for Acquisition/Subscription Address
300	A	Physical Description
315	O	Frequency (serial)
340	A	Medium
350	O	Price
351	O	Organization and Arrangement
362	A	Date of Publication and/or Volume Designation
400	A	Series Statement/Added Entry – Personal Name
410	A	Series Statement/Added Entry – Corporate Name
411	A	Series Statement/Added Entry – Meeting Name
440	A	Series Statement – Title
490	A	Series Statement
500	O	General Note
501	O	With Note
502	O	Dissertation Note
503	O	Bibliographic History Note (see 581)
504	O	Bibliography Note
505	O	Formatted Contents Note
506	O	Restrictions on Access
510	O	Citations/References Note (Cited in?)
516	O	Type of File or Data Note
520	O	Summary, Abstract, Annotation, Scope Note
521	O	Target Audience (level)
522	O	Geographic Coverage
523	O	Chronological Coverage of Data Collection (see 045)
524	O	Preferred Citation of Described Materials
530	O	Additional Physical Form Available
535	O	Location of Originals/Duplicates
536	O	Sponsoring/Funding Information
537	O	Source of Data
538	O	Technical Details
556	O	Information about Documentation
565	O	Case File Characteristics

580	A	Linking Entry Complexity
581	O	Primary Publications (User Guides – see 503)
582	O	Related Computer Files (MRDFs)
583	O	Action Taken (Processing Note)
59x	O	Local Notes (see holdings area?)
600	A	Subject Added Entry – Personal Name/Title
610	A	Subject Added Entry – Corporate Name/Title
611	A	Subject Added Entry – Meeting Name
630	A	Subject Added Entry – Uniform Title
650	A	Subject Added Entry – Topical Headings
651	A	Subject Added Entry – Geographic Name
653	A	Subject Added Entry – uncontrolled index term
69x	O	Local Subject Access Fields
700	A	Added entry – Personal Name/Title
710	A	Added entry – Corporate Name/Title
711	A	Added entry – Conference Title
730	A	Added entry – Uniform Title
740	A	Added entry – Variant Title (traced differently)
753	O	Technical Details Access to Computer Files
755	O	Added Entry – Physical Characteristics Access
773	A	Host Item Entry
800	A	Series Added Entry – Personal Name
810	A	Series Added Entry – Corporate Name
811	A	Series Added Entry – Meeting Name
830	A	Series Added Entry – Uniform Title